Published online 27 November 2007 *Nucleic Acids Research, 2008, Vol. 36, Database issue* ***D689–D694***
doi:10.1093/nar/gkm940

EndoNet: an information resource about regulatory networks of cell-to-cell communication†

Jürgen Dönitz[1], Björn Goemann[1], Muriel Lizé[1], Holger Michael[1], Nicole Sasse[1], Edgar Wingender[1,2] and Anatolij P. Potapov[1,*]

[1]Department of Bioinformatics, University of Göttingen/Medical School, D-37077 Göttingen and [2]BIOBASE GmbH, D-38304 Wolfenbüttel, Germany

Received September 14, 2007; Revised October 8, 2007; Accepted October 11, 2007

ABSTRACT

EndoNet is an information resource about intercellular regulatory communication. It provides information about hormones, hormone receptors, the sources (i.e. cells, tissues and organs) where the hormones are synthesized and secreted, and where the respective receptors are expressed. The database focuses on the regulatory relations between them. An elementary communication is displayed as a causal link from a cell that secretes a particular hormone to those cells which express the corresponding hormone receptor and respond to the hormone. Whenever expression, synthesis and/or secretion of another hormone are part of this response, it renders the corresponding cell an internal node of the resulting network. This intercellular communication network coordinates the function of different organs. Therefore, the database covers the hierarchy of cellular organization of tissues and organs as it has been modeled in the Cytomer ontology, which has now been directly embedded into EndoNet. The user can query the database; the results can be used to visualize the intercellular information flow. A newly implemented hormone classification enables to browse the database and may be used as alternative entry point. EndoNet is accessible at: http://endonet.bioinf.med.uni-goettingen.de/

INTRODUCTION

Currently, most efforts in systems biology are made on various aspects of intracellular regulation. Great progress has been achieved in understanding the organization and functioning of various regulatory pathways and networks. Several databases have been created to collect and organize the corresponding information on these special issues. Representative examples of such databases and knowledge bases containing information about intracellular regulatory interactions are KEGG (1,2), BIND (3), DIP (4), aMaze (5), EcoCyc (6), GeneNet (7), Reactom (8), TRANSPATH (9) and TRANSFAC (10). Altogether, they form the necessary basis for modeling and simulation of large intracellular networks. Compared to this, large-scale intercellular molecular networking, which is of key importance for multicellular eukaryotes, has been largely neglected in these modeling attempts.

In multicellular organisms, the functional activities of various cell types, tissues and organs need to be coordinated to provide proper functioning of the whole system (11). A great body of information on hormones, growth factors, cytokines and other messengers involved in intercellular cross talking has been published (12–15). However, it is mainly about elementary cell–cell interactions, although some attempts to integrate such interactions into more complex regulatory pathways can be found in the literature as well (12–15). Cell-to-cell regulatory interactions and networks of intercellular regulatory communication still remain challenges in the field of systems biology.

The need to combine the available information on cell-to-cell interaction was realized recently when the EndoNet database was developed (16). Many of such intercellular regulatory pathways display the property of a cascade in which the initial signal is multiplied over several steps. Moreover, such networks typically include multiple regulatory circuits. EndoNet aims at modeling these network topologies and at bridging the existing gap between known genotypes and their molecular and clinical phenotypes, thus allowing the utilization of EndoNet in medical research. During the last years, the EndoNet database was subject to considerable expansion of its contents and improvements of its data organization and presentation. Among the most prominent new features, the implementation of a hormone classification for

*To whom correspondence should be addressed. Tel: +49 551 39 14917; Fax: +49 551 39 14914; Email: apo@bioinf.med.uni-goettingen.de

†If the EndoNet database is used in any published research, the authors wish that this paper be cited.

browsing and the full embedding of our Cytomer ontology on cells, tissues, organs and their developmental stages (17,18) will be described in this paper.

HORMONES AND THEIR CLASSIFICATION IN ENDONET

Historically, hormones were defined as substances produced and secreted by special endocrine glands. However, it is commonly accepted now that nearly all tissue types are capable of producing molecular signals that are informative for their direct neighbors or remote counterparts and alter their growth, function or metabolism (13–15). Therefore, with the term 'hormones' we refer to a broad spectrum of biomolecules, which are used for cell-to-cell communication and help to coordinate complex biological functions such as growth and differentiation by directing gene expression, secretory and metabolic activities of various cells, tissues and organs. This includes 'classical hormones', as well as all cytokines, growth factors and many others. Currently there are 637 hormones (in the broadest sense) in the EndoNet database (Table 1). This group is highly heterogeneous in regard to the properties of such molecules, chemical structure and the type of action.

By considering the origin of hormones and their relations to the genome, all molecular messengers in EndoNet have been divided into two superfamilies. The first superfamily represents those hormones which are not directly encoded by genes. Such 'not genome-encoded' hormones are typically small molecules that are synthesized by a set of metabolic pathways. Most of them bind to intracellularly located receptors. In general, intracellular regulatory pathways that are activated by the hormones of this superfamily are relatively short and simple.

The superfamily of 'not genome-encoded' hormones includes several families, many of which are further subdivided in subfamilies and sub-subfamilies. This may be illustrated with an example of the 'androgens' sub-subfamily all members of which (5α-androstanediol, androstenedione, dehydro-3-epiandrosterone, DHT and testosterone) belong to the subfamily of steroids, which in turn is part of the sterol lipids family (Figure 1).

Table 1. Contents of EndoNet in comparison with the previous release

Components	Number of entries (17 October 2005)	Number of entries (15 September 2007)
Molecules		
Hormone	109	637
Receptors	117	500
Cellular sources		
Cells/tissues	112	314
Relations		
Hormone—receptor	149	861
Donor cell—hormone	184	1920
Receptor—acceptor cell	292	1555
Information sources		
References	264	1926

The second superfamily includes peptide and protein hormones that are encoded in the genome by individual or multiple genes, in case that the messengers are multi-subunit proteins. This superfamily is much larger and more diverse. It consists of multiple families, which are further divided into subfamilies and sub-subfamilies. Thus, the family of cytokines includes multiple subfamilies and one of them, chemokines, consists of sub-subfamilies of C chemokines, CC chemokines and CXC chemokines. To become active, most of the peptide/protein hormones require more or less extensive posttranslational processing, e.g. leader peptide cleavage or cutting of a precursor polypeptide into a series of shorter poly- and/or oligopeptides. In all cases we are aware of, genome-encoded hormones initiate the corresponding cellular responses by binding to specific membrane-associated receptors. These receptors usually trigger more complex intracellular signaling cascades towards the nucleus (11,13).

To search hormone superfamilies, the user may browse the 'hormone classification' that can be found at the 'Search Item' page of the EndoNet user interface. All primary entities are linked to a detail page, thereby enabling easy access to the available descriptive information.

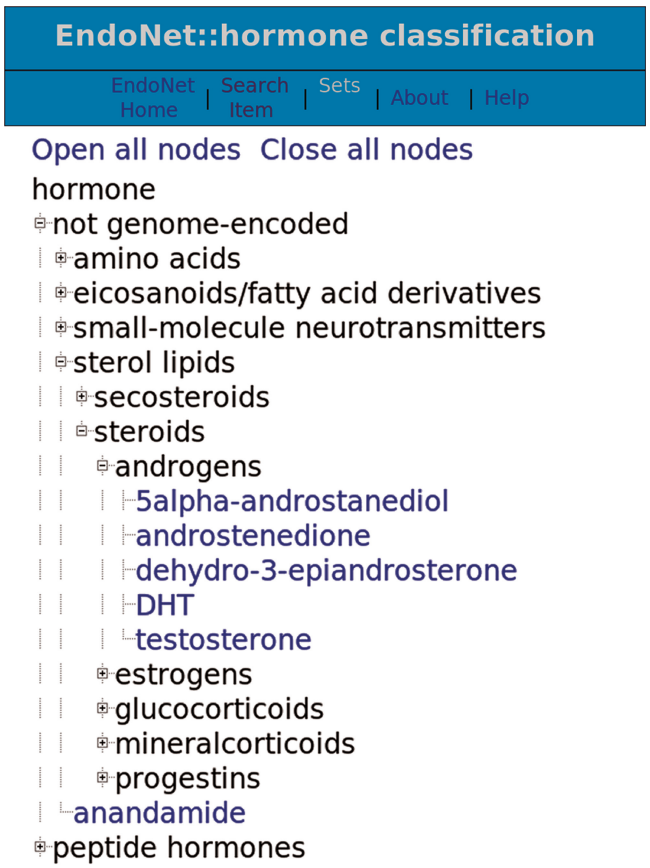

Figure 1. Classification of hormones in the EndoNet database. The fragment related to the subsubfamily of androgens is shown. The basic entities (e.g. 5α-androstanediol, androstenedione) are clickable thereby enabling easy access to the contents of the database. Accordingly, the classification can be used for making queries.

We are aware that the hormone classification provided is just a starting point; in particular most parts of the genome-encoded family are very flat and just listed alphabetically. Work is in progress that identifies proper criteria for a structure-driven classification which may reflect as many functional features as possible.

THE CYTOMER ONTOLOGY

Cytomer is an ontology of anatomical structures (17,18). In an ontology, the entities described are represented as classes. These classes form a hierarchy and are connected to each other through the relation isA (19).

The Cytomer ontology comprises two major classes, AnatomicalEntity and AnatomicalAbstractConcept. The former combines all anatomical structures, whereas the latter contains notional abstractions like Species or DevelopmentalStage. The upper structure of the ontology is shown in Figure 2. Although this structure provides for the inclusion of different species, most of the information present in Cytomer relates to human.

The components of the human body are subsumed in the class AnatomicalEntity, which contains fluids (class Humour), lacunae (AnatomicalCavity) and the diverse body parts (AnatomicalBuildingBlock). The latter class is divided into several subclasses, including but not restricted to Organ, OrganPart, and Cell. These classes are organized as a directed acyclic graph (19). The relations between the classes are modeled either as class–subclass relations (isA) or as properties of the classes (hasPart). In Cytomer, these relation types are modeled in a reciprocal way: thus isA is equivalent to hasSubclass, and hasPart is equivalent to isPartOf. Figure 2 shows the relations between the alpha cells of the Langerhans islets

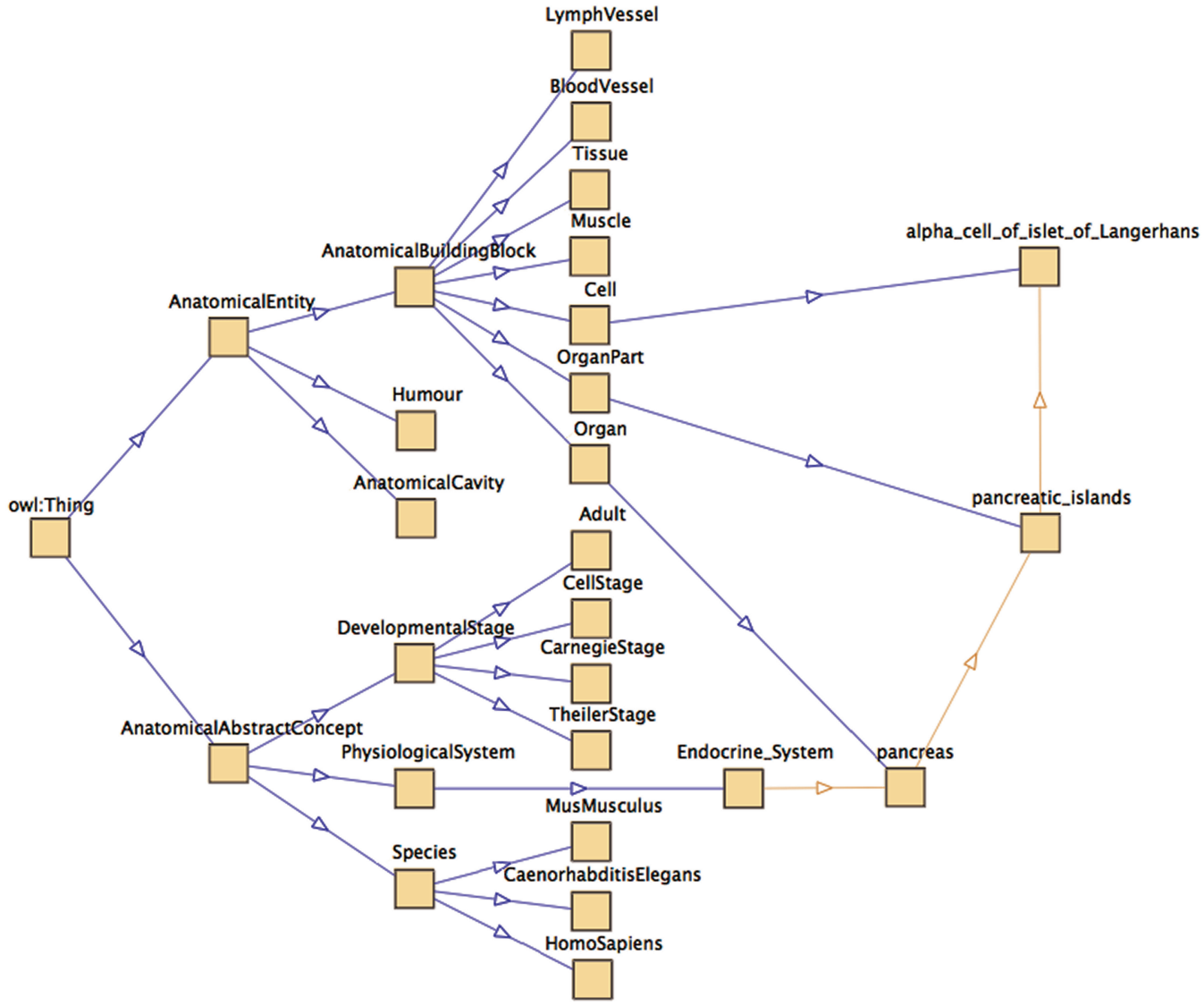

Figure 2. Visualization of Cytomer. Shown are the upper structure of the ontology (left) and a part of the lower hierarchy, exemplifying the partOf/hasPart relationship (right). Boxes denote classes, blue arrows indicate the hasSubclass relationship and orange arrows indicate the hasPart connection.

and its related classes as an example. The diverse body components are also related to several physiological systems as well as to the different developmental stages (Carnegie stages in case of human). The latter connection models the different temporal occurrences of the entities described in the successive phases of the embryonic development.

CYTOMER CONNECTOR

Cytomer is stored and modeled with the Web Ontology Language (OWL). This allows to access Cytomer via Application Programming Interfaces (API) such as the Jena-API and the Protégé-API. Both are generic toolkits designed to handle any ontology based on OWL. Because of the need for functionality that is specific to an anatomical ontology, we have developed the Cytomer connector, which uses the Protégé-API and provides simple access to Cytomer (Figure 3).

In contrast to many other information resources which are accessible in EndoNet via hyperlinks, Cytomer is directly embedded into the EndoNet user interface (Figure 4). With the Cytomer viewer, the user can explore the whole ontology by browsing through the graph of interest and analyze all connected anatomical properties. In spite of the integrated view on Cytomer and EndoNet, from the technical point of view, both systems are technically independent and separately maintained by experts in the respective fields.

CONTENTS

All entries in EndoNet have been created by manual annotation and are extracted from original publications, monographs (12–15) and the linked databases. As a tool for systems biology EndoNet focuses on the macroscopic view of the information flow. For additional detailed information on single components and events, external data sources are referenced. Data on genome-encoded entities, namely peptide/protein hormones and receptors, include references to external resources such as Swiss-Prot [(20), http://www.expasy.org/sprot/], TRANSPATH (9) and HumanPSD™ (21) and NCBI (http://www.ncbi.nlm.nih.gov/). Data on not genome-encoded hormones include references to the databases KEGG (1,2), Lipid Maps [(22), http://www.lipidmaps.org/] and Lipidbank (http://lipidbank.jp/).

The annotation policy for EndoNet is to record the information as precisely as possible without generalization. The endocrine substances and receptors are assigned to cells, tissues and organs as specifically as possible. All genome-encoded molecules are represented in their active, i. e. processed or multisubunit form, although the links to external databases such as Swiss-Prot and HumanPSD usually refer to the corresponding precursor or subunit polypeptides.

In comparison with the previous release published in 2006 (16), the number of entries for hormones, receptors and cells/tissues/organs and in particular the number of connections between these entities has significantly increased as shown in Table 1. In the publicly accessible version, there are no orphan entries for hormones and receptors: that is, all of them are linked to the corresponding receptor or ligand entities, respectively, thereby increasing the size and the edge density of the whole network.

DATA ACCESS AND VISUALIZATION

The access to EndoNet is provided through a servlet-based web interface. The contents of EndoNet can be browsed or searched. Searching for a specific hormone or receptor can be done by name or synonym or a substring thereof. To facilitate retrieval of a hormone when just the

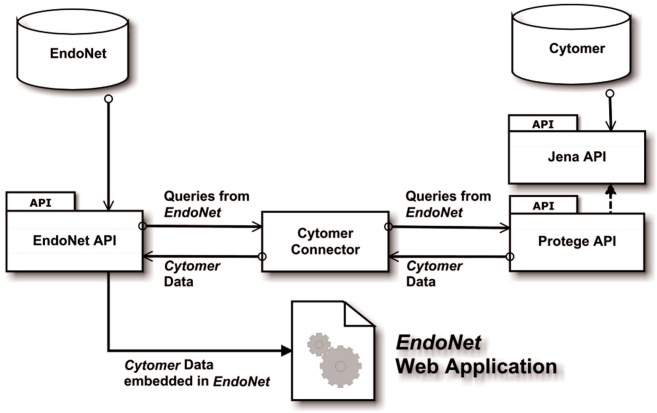

Figure 3. Schematic view of the information flow between the components of EndoNet and Cytomer.

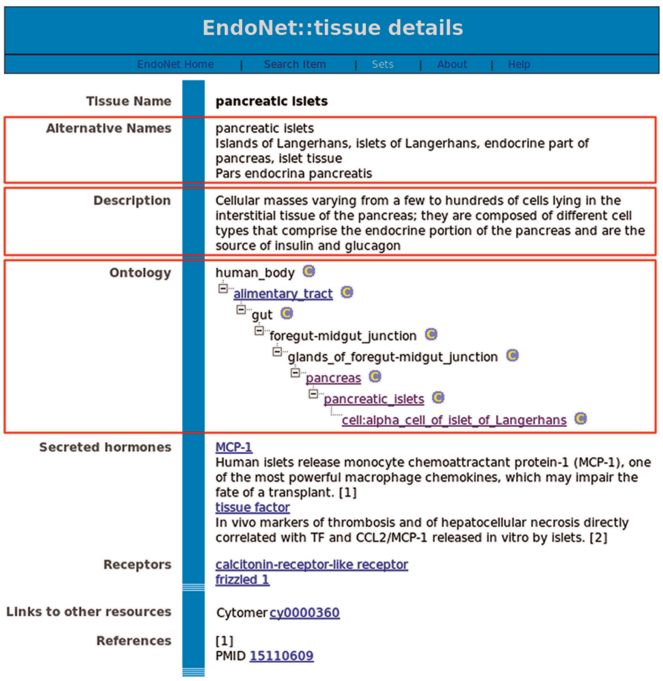

Figure 4. Integrated view on EndoNet and Cytomer entities. The result of a query for pancreatic islets is displayed as an illustrative example. The parts within red boxes relate to the Cytomer ontology.

family it belongs to but not the exact name is known, the newly implemented hormone classification offers an alternative search strategy (Figure 1). For cells, tissues and organs the search algorithm includes the medical name and synonyms from Cytomer (Figure 4). Several queries can be combined to build a network.

The result of a query is a list of matching entities which can be added to a user-defined set to be combined with the results of further searches. From this set a network can be built using its items as starting points for up- and downstream searches. Such a graph is displayed using Graphviz (8). Hormones and receptors are shown as vertices grouped together into subgraphs that represent the cells/tissues/organs where they are secreted from or expressed in, respectively. 'Activation' and 'inhibition' actions are shown as different types of edges.

All objects in EndoNet whether displayed textually or graphically are linked to a corresponding detail page. Here all relevant information from all data sources is combined. This comprises alternative names, lists of connected EndoNet objects, links to external databases and to the original literature. Some well-established endocrine pathways are offered in the form of 'predefined sets'. When selected, such sets can be used for obtaining a quick overview or as a starting point for more complex queries.

DISCUSSION

By now EndoNet has reached a high level of coverage of existing knowledge about intercellular regulatory communication in humans. The amount of data has increased significantly since the first publication. The usability was improved by adding meta-information to the core data, enabling the user to browse the anatomical hierarchy provided by Cytomer and the newly introduced hormone classification. To unburden the user from the task to combine the relevant information from two independent resources, we did not just link Cytomer to EndoNet, but rather truly embedded the ontology into the EndoNet interface.

Besides being an encyclopedia on hormones, receptors, various cells, tissues, organs and direct regulatory links between them, EndoNet can serve for modeling the integral organization of large intercellular regulatory networks. As mentioned, many entries, i.e. hormones and receptors, in EndoNet have links to external information resources including TRANSPATH (9). The latter is one of the few databases providing information about signal transduction and gene expression in higher eukaryotes and in particular in mammals. We see a great advantage in combining these complementary databases into a synergistic system, which is capable of integrating multiple intracellular and intercellular regulatory networks. This system thus allows for making an integral analysis of human regulatory pathways by means of significantly extended level of molecular details (23,24). In addition, we plan to expand the EndoNet view onto rodent components as soon as the human network has reached a more complete status.

EndoNet represents an important contribution to systems biology approaches by exceeding the previous emphasis on intracellular objects and processes. As far as we know, EndoNet is the first and yet the only system capable of modeling large-scale intercellular regulatory networks in humans. The purpose of this database is to provide a useful resource for biomedical research in its different facets like pharmacologic research, diagnostics or therapeutic aspects.

Finally, EndoNet aims at bridging the existing gap between known genotypes and their molecular and clinical phenotypes, thus allowing utilization of EndoNet in medical research. Considering the high complexity of this task, we would like to emphasize that EndoNet as well as Cytomer are ongoing projects.

Availability

EndoNet is freely available for non-commercial use under http://endonet.bioinf.med.uni-goettingen.de/.

ACKNOWLEDGEMENTS

This work was funded in part by 'Forschungs- und Berufungspool des Niedersächsischen Ministeriums für Wissenschaft und Kultur' (Kap. 06 08 TG 74) and by the Federal Ministry of Education, Research and Technology (BMBF) as part of the National Genome Research Network (NGFN, grant NGFN2 – Biomed – SMP – Bioinformatik, FKZ: 01GR0-480). The authors gratefully acknowledge BIOBASE GmbH for providing free access to linked entities in HumanPSD and thank Torsten Schoeps for expert technical assistance. We are indebted to Charles Elfe (BIOBASE Corp.) for providing us with the links to HumanPSD entries. We are also grateful to the anonymous reviewers for their critical comments and recommendations. TRANSPATH, TRANSFAC and Cytomer are registered trademarks; HumanPSD is a trademark of BIOBASE GmbH, Wolfenbüttel, Germany. Funding to pay the Open Access publication charges for this article was provided by the University of Göttingen/ Medical School, Department of Bioinformatics.

Conflict of interest statement. None declared.

REFERENCES

1. Kanehisa,M., Goto,S., Kawashima,S., Okuno,Y. and Hattori,M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **32**, D277–D280.
2. Kanehisa,M., Goto,S., Hattori,M., Aoki-Kinoshita,K.F., Itoh,M., Kawashima,S, Katayama,T., Araki,M. and Hirakawa,M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354–D357.
3. Bader,G.D., Betel,D. and Hogue,C.W. (2003) BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.*, **31**, 248–250.
4. Salwinski,L., Miller,C.S., Smith,A.J., Pettit,F.K., Bowie,J.U. and Eisenberg,D. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.*, **32**, D449–D451.
5. Lemer,C., Antezana,E., Couche,F., Fays,F., Santolaria,X., Janky,P., Deville,Y., Richelle,J. and Wodak,S.J. (2004) The aMAZE LightBench: a web interface to a relational database of cellular processes. *Nucleic Acids Res.*, **32**, D443–D448.
6. Keseler,I.M., Collado-Vides,J., Gama-Castro,S., Ingraham,J., Paley,S., Paulsen,I.T., Peralta-Gil,M. and Karp,P.D. (2005)

EcoCyc: a comprehensive database resource for Escherichia coli. *Nucleic Acids Res.*, **33**, D334–D337.
7. Ananko,E.A., Podkolodny,N.L., Stepanenko,I.L., Podkolodnaya,O.A., Rasskazov,D.A., Miginsky,D.S., Likhoshvai,V.A., Ratushny,A.V., Podkolodnaya,N.N. *et al.* (2005) GeneNet in 2005. *Nucleic Acids Res.*, **33**, D425–D427.
8. Joshi-Tope,G., Gillespie,M., Vastrik,I., D'Eustachio,P., Schmidt,E., de Bono,B., Jassal,B., Gopinath,G.R., Wu,G.R. *et al.* (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, **33**, D428–D432.
9. Krull,M., Pistor,S., Voss,N., Kel,A., Reuter,I., Kronenberg,D., Michael,H., Schwarzer,K., Potapov,A. *et al.* (2006) TRANSPATH®: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Res.*, **34**, D546–D551.
10. Matys,V., Kel-Margoulis,O.V., Fricke,E., Liebich,I., Land,S., Barre-Dirrie,A., Reuter,I., Chekmenev,D., Krull,M. *et al.* (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.
11. Lodish,H., Berk,A., Zipursky,S.L., Matsudaira,P., Baltimore,D. and Darnell,J. (2000) *Molecular Cell Biology*. Media Connected, W.H. Freeman and Company, NY.
12. Kronenberg,H., Melmed,S., Larsen,P.R. and Polonsky,K. (2003) Principles of endocrinology. In Larsen,P.R., Kronenberg,H.M, Melmed,S. and Polonsky,K.S. (eds), *Williams Textbook of Endocrinology*. Saunders, Elsevier Science, Philadelphia, pp. 1–9.
13. Nussey,S.S. and Whitehead,S.A. (2001) *Endocrinology: An Integral Approach*. BIOS Scientific Publishers Ltd., Oxford.
14. Martini,L. (ed) (2004) *Encyclopedia of Endocrine Diseases*, Vol. 1–4. Elsevier Inc., Academic Press, Amsterdam.
15. Norris,D.O. (2007) *Vertebrate Endocrinology,* 4th edn. Elsevier Inc., Academic Press, Amsterdam.
16. Potapov,A., Liebich,I., Dönitz,J., Schwarzer,K., Sasse,N., Schoeps,T., Crass,T. and Wingender,E. (2006) EndoNet: an information resource about endocrine networks. *Nucleic Acids Res.*, **34**, 545.
17. Heinemeyer,T., Chen,X., Karas,H., Kel,A.E., Kel,O.V., Liebich,I., Meinhardt,T., Reuter,I., Schacherer,F. *et al.* (1999) Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms. *Nucleic Acids Res.*, **27**, 318–322.
18. Michael,H., Chen,X., Fricke,E., Haubrock,M., Ricanek,R. and Wingender,E. (2004) Deriving an ontology for human gene expression sources from the CYTOMER® database on human organs and cell types. *In Silico Biol.*, **5**, 0007.
19. Bard,J.B. and Rhee,S.Y. (2004) Ontologies in biology: design, applications and future challenges. *Nat. Rev. Genet.*, **5**, 213–222.
20. Wu,C.H., Apweiler,R., Bairoch,A., Natale,D.A., Barker,W.C, Boeckmann,B., Ferro,S., Gasteiger,E., Huang,H. *et al.* (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.*, **34**, D187–D191.
21. Hodges,P.E., Carrico,P.M., Hogan,J.D., O'Neill,K.E., Owen,J.J., Mangan,M., Davis,B.P., Brooks,J.E. and Garrels,J.I. (2002) Annotating the human proteome: the Human Proteome Survey Database (HumanPSD™) and an in-depth target database for G protein-coupled receptors (GPCR-PD™) from Incyte Genomics. *Nucleic Acids Res.*, **30**, 137–141.
22. Fahy,E., Sud,M., Cotter,D. and Subramaniam,S. (2007) LIPID MAPS online tools for lipid research. *Nucleic Acids Res.*, **35**, W606–W612.
23. Wingender,E., Crass,T., Hogan,J.D, Kel,A.E., Kel-Margoulis,O.V. and Potapov,A.P. (2007) Integrative content-driven concepts for bioinformatics "beyond the cell". *J. Biosci.*, **32**, 169–180.
24. Wingender,E., Hogan,J., Schacherer,F., Potapov,A.P. and Kel-Margoulis,O. (2007) Integrating pathway data for systems pathology. *In Silico Biol.*, **7**, 03.